DOCUMENT RESUME

ED 152 658 SO 010 808

AUTHOR Pfau, Richard H.
TITLE Factors Affecting the Validity of Cross-National Comparisons of Classroom Behaviors.
INSTITUTION Connecticut Univ., Storrs. Inst. of Public Service.
PUB DATE Feb 78
NOTE 40p.; Paper presented at Annual Meeting of the Society for Cross-Cultural Research (7th, New Haven, Connecticut, February 23-26, 1978);

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Classroom Communication; *Classroom Observation Techniques; *Classroom Research; Comparative Analysis; Comparative Education; *Cross Cultural Studies; Feasibility Studies; Foreign Countries; Interaction Process Analysis; Measurement Instruments; Reliability; *Research Problems; Sampling; Social Systems; *Student Behavior; Teacher Behavior; *Validity
IDENTIFIERS Activity Categories Instrument; Flanders System of Interaction Analysis; Nepal

ABSTRACT

Written to assist future researchers in conducting cross-national studies of classroom behaviors, this study indicates a number of factors which may reduce the validity of comparisons made. It is based on an investigation which explored the feasibility of using the Activities Categories Instrument and the Flanders Interaction Analysis Categories as a basis for making valid comparisons between classroom behaviors in the United States and Nepal. These instruments consist of categories of behavior, recorded by an observer, which can be used to describe teacher and student talk. Some of the 15 factors discussed include: teachers and students may behave differently than normal, given the presence of an observer; since category systems make use of time units to standardize the frequency or pacing of an observer's recording, a certain amount of distortion takes place because classroom events may break on or between the arbitrary time boundaries; researchers may encounter problems associated with similar but functionally different behaviors, and functionally equivalent but different behaviors; and the failure of personnel to follow a project's planned methodology either unconsciously or as a result of dishonesty. It is concluded that further research is needed to explore the strengths and weaknesses associated with category system usage. (Author/JK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

APR 14 1978

ED152658

FACTORS AFFECTING THE VALIDITY OF CROSS-NATIONAL COMPARISONS OF CLASSROOM BEHAVIORS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard H. Pfau

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Richard H. Pfau
Institute of Public Service
The University of Connecticut
Storrs, CT 06268

SP 010 808

Paper presented at the seventh annual meeting
of the Society for Cross-Cultural Research
New Haven, Connecticut, 23-26 February 1978

Introduction

At the present time, social scientists have only vague ideas about what occurs in the classrooms of most countries of the world, and how classroom behaviors vary from one social system to another.

This paper has been written to assist future researchers in conducting cross-national studies of classroom behaviors by indicating a number of factors which may serve to reduce the validity of comparisons made. It is based upon insights gained during a study in which this investigator (a) explored the feasibility of using category systems as a basis for making objective and valid cross-national comparisons of classroom behaviors, and (b) surveyed and described a number of classroom behaviors occurring within the Kingdom of Nepal and determined how those behaviors differed from behaviors occurring in U.S. classrooms (Pfau et al., 1975; Pfau, 1977a, 1977b).

The review which follows focuses upon factors associated with the use of direct observation techniques in general, and several category systems in particular, for the use of category systems seems to provide a potential for making precise and valid comparisons not equaled by other techniques used to measure and describe classroom behaviors (Pfau, 1976). Factors to be discussed are:

A. Sampling Concerns

1. Sample Drop Out
2. Observer Effects Upon Classroom Behaviors
3. Population Comparability
4. Quality of Sampling Frames
5. Sample Sizes and Representativeness
6. Cycles in Classroom Behaviors

B. Encoding Concerns

1. Intra-Investigation Reliability
2. Observer Drift in Reliability
3. Time Unit Distortion
4. Use of Different Encoding Tempos
5. Encoding Subjectivity

C. Decoding Observer Records

1. Equivalence of Behavior Indices
2. Problems of Context and Meaning
3. The Sequential Analysis of Behaviors

D. Failure to Follow Project Protocol

The FIAC and ACI Category Systems

During the study of Nepalese and U.S. classroom behaviors, behavioral measures compared were based upon the use of two category systems. These category systems were the Flanders Interaction Analysis Categories and the Activity Categories Instrument.

The Activity Categories Instrument (ACI), developed by Caldwell (1967, 1968), is shown in Table 1. As can be seen, this instrument consists of 11 categories of behavior which describe types of activities occurring in science classes. An observer using this instrument makes a record of classroom activities with the aid of a stopwatch. Every 5 seconds the observer notes a single number associated

TABLE 1

ACTIVITY CATEGORIES INSTRUMENT

1. LABORATORY EXPERIENCES: OPENENDED. Students are presented a problem to be solved by experimentation. The procedure may or may not be given. They are required to make observations and analyze or interpret their findings.

2. LABORATORY EXPERIENCES: STRUCTURED. Students are presented a laboratory experiment with a structured procedure. They are not required to analyze or interpret their data. They are asked to make observations.

3. GROUP PROJECTS. One or more groups of students are working on a science project during the class period. Some may work individually. (not written projects)

4. STUDENT DEMONSTRATIONS. A student or group of students demonstrate a science experiment or project which they have prepared. (oral report on science project would be included)

5. STUDENT LIBRARY RESEARCH, REPORTING, ETC. (a) A student or group of students give an oral report they have prepared based on reference material. (b) The class works with reference materials for purposes of writing or making reports.

6. STUDENT SPEAKING. The student contributes verbally by asking a question, answering a question or simply volunteering information.

7. TEACHER QUESTIONING. Students are asked a question by the teacher.

8. WORKBOOK WORK. Students work in class on workbooks, homework, questions from text, art-type work, etc.

9. TEACHER DEMONSTRATIONS. The teacher presents material by film, filmstrip, record, TV, radio, demonstration, etc.

10. LECTURE. The teacher reads aloud, expresses his views, gives directions, makes an assignment or asks rhetorical questions. Students are expected to listen. They may interrupt only when they do not understand. Student reading in the text is also included.

11. GENERAL HAVOC. The class may be cleaning up, settling down or doing nothing. In general, this category should be used sparingly.

(Caldwell, 1967, 1968)

with one of the ACI categories of behavior. The activity which occurs during each 5 second interval determines which category number is recorded. Thus, if the teacher was asking a question, the observer records a 7, (teacher questioning). If a student response to the question occupied the next 5 second interval, the number 6 (student speaking) would be recorded. In this way a series of numbers is recorded that indicates which of the 11 categories of behavior occurred during a period of observation, and the relative frequency of their occurrence.

Table 2 contains a description of the Flanders Interaction Analysis Categories (FIAC), perhaps the most frequently used category system yet developed (Dunkin & Biddle, 1974, p. 100). This instrument consists of 10 categories of behavior, which can be used to describe teacher and student talk occurring in the classes of any subject-matter area. An observer using this instrument records a category number representing the behavior observed about every 3 seconds. If more than one behavior of concern occurs during a 3 second interval, then all categories used in that interval are recorded (Amidon & Flanders, 1971, p. 27). Observers are sufficiently trained to do this without the aid of watches. The numbers are recorded in sequence on an observation sheet. A recording of . . . 3,5,4,8,2,4, . . . , for example would indicate that the teacher lectured, asked a question which was followed by a student response, praise or encouragement was then

TABLE 2

FLANDERS INTERACTION ANALYSIS CATEGORIES

TEACHER TALK	1. ACCEPTS FEELING. Accepts and clarifies an attitude or the feeling tone of a pupil in a nonthreatening manner. Feelings may be positive or negative. Predicting and recalling feelings are included. 2. PRAISES OR ENCOURAGES. Praises or encourages pupil action or behavior. Jokes that release tension, but not at the expense of another individual; nodding head, or saying "Um hm?" or "go on" are included. 3. ACCEPTS OR USES IDEAS OF PUPILS. Clarifying, building, or developing ideas suggested by a pupil. Teacher extensions of pupil ideas are included but as the teacher brings more of his own ideas into play, shift to category five. 4. ASKS QUESTIONS. Asking a question about content or procedure, based on teacher ideas, with the intent that a pupil will answer. 5. LECTURING. Giving facts or opinions about content or procedures; expressing his own ideas, giving his own explanation, or citing an authority other than a pupil. 6. GIVING DIRECTIONS. Directions, commands, or orders to which a pupil is expected to comply. 7. CRITICIZING OR JUSTIFYING AUTHORITY. Statements intended to change pupil behavior from nonacceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what he is doing; extreme self-reference.
STUDENT TALK	8. PUPIL TALK--RESPONSE. Talk by pupils in response to teacher. Teacher initiates the contact or solicits pupil statement or structures the situation. Freedom to express own ideas is limited. 9. PUPIL TALK--INITIATION. Talk by pupils which they initiate. Expressing own ideas; initiating a new topic; freedom to develop opinions and a line of thought, like asking thoughtful questions; going beyond the existing structure.
	10. SILENCE OR CONFUSION. Pauses, short periods of silence and periods of confusion in which communication cannot be understood by the observer.

Note. There is no scale implied by the numbers indicated. Each number is classificatory; it designates a particular kind of communication event.

(Flanders, 1970, p. 34)

given by the teacher, and he asked another question.

Observers using category systems such as these, are usually trained until their recordings agree highly with those made by experts in the use of the instrument. This high agreement helps ensure that observations are made in a standardized way. Also, the fact that high agreement can be reached indicates that the process of assigning numbers to describe observations is unambiguous.

Once recordings are made, they can be analyzed in a number of ways. For example, frequency counts can be made to determine how many times each category number was recorded during the period of observation. Based upon these, the percentage of time during which each category of behavior occurred can be estimated. If observations are made of a number of classes, means or medians can then be calculated and compared with those of other samples.

Meanings of Validity

Validity may be thought of as referring to the appropriateness of inferences made based upon measurement scores (Standards for Educational & Psychological Tests, 1974, p. 25). In a comparative context, validity indicates that we are measuring in each system under consideration what we intended to measure (Przeworski & Teune, 1970, p. 103).

In general, the validity of classroom measurements seems to depend upon the fulfillment of at least three con-

ditions. First, a representative sample of the behaviors to be measured must be observed. Secondly, an accurate record of the behaviors observed must be made. And thirdly, the records must be scored and interpreted so as to faithfully reflect similarities and differences in behavior, and faithfully recreate elements of the original situations in their proper perspectives (Medley & Mitzel, 1963, p. 250; Flanders, 1970, p. 87). Factors concerning each of these conditions are reviewed in the following sections, and related to the study of Nepalese and U.S. classroom behaviors conducted by this investigator. Although judgments of the validity of measurements and comparisons made during that study were high (Pfau, 1977a; 1977b, pp. 213-218), the factors to be discussed seem worthy of consideration by future researchers and are reviewed here for that reason.

Sampling Concerns

Sample Drop Out

Even if sound sampling strategies are used in an observational study, the possibility exists that some school administrators or teachers will not allow classes to be observed. Flanders found that at least 10% and sometimes 50% of U.S. teachers selected for observation, refused to permit their classes to be observed (Flanders, 1970, p. 386). This type of drop out can make a sample

observed less representative.

There are no clear-cut solutions to this problem. Substituting a teacher who is willing to cooperate for one who is not willing, by no means ensures representativeness (Flanders, 1970, p. 386). Also, if rigorous sampling procedures are used, it seems likely that some teachers will be selected who would rather not be observed. These teachers, if they are forced to participate, may alter their behavior more than volunteers. Thus, more rigorous procedures may produce less representative classroom interaction data (Flanders, 1969, vol. 1, p. 5).

When considered with a cross-national perspective, the influence of nonparticipants upon the validity of comparisons made between countries is complicated by the likelihood that different types of drop out will occur in the culturally different populations being studied. In this regard, it might be pointed out that, although teacher drop out is apparently a problem encountered in the U.S.A., it was not a problem encountered within Nepal. No Nepalese teachers or principals refused to permit observations of classes to be made.

Observer Effects upon Classroom Behaviors

Another sampling problem is derived from the fact that an observer's presence may affect the classroom behaviors that he measures. That is, teachers and students may behave somewhat differently than they normally would

under the gaze of an observer. After all, teaching is usually a fairly private activity, not conducted in the presence of adults who have intruded into the classroom (Dunkin & Biddle, 1974, p. 58).

A number of persons, however, feel that observer influence need not be a great problem. Medley and Mitzel, for example, indicated that if teacher and school anonymity is guaranteed, if teachers are ensured that their behaviors are not being evaluated, and if teachers have an idea of why the observations are being made, then observer effects can be minimized (1963, pp. 247-248, & 306-307). In a similar vein, Kerlinger felt that

> observers seem to have little effect on the situations they observe. Individuals and groups seem to adapt rather quickly to an observer's presence and to act as they would usually act. This does not mean that the observer cannot have an effect. It means that if the observer takes care to be unobtrusive and not to give the people observed the feeling that judgments are being made, then the observer as an influential stimulus is mostly nullified. (Kerlinger, 1973, p. 539)

A number of researchers who have conducted observational studies, also seem to have felt that teachers and students quickly forgot about the observer's presence and behaved as they normally would (see for example R. S. Adams & Biddle, 1970, pp. 22-25; and Snider, 1966, pp. 79-82).

However, Samph (1968; 1976) in a study of 10 female teachers using the FIAC found that teacher verbal behavior varied as a function of an observer's presence or absence in the classroom. For example, teachers used more praise,

more acceptance of student ideas, and less criticism when they knew that they were being observed. Nuthall and Church (1972) also discussed a study which "found that the presence of TV cameras resulted in teachers speaking more often and more slowly, and pupils speaking less often" (p. 493). They noted, though, that these effects disappeared after about 2 days.

Thus, to summarize, it appears that observers can have an effect upon classroom behaviors observed. In the U.S.A. at least (and apparently also in Nepal), these effects can be minimized by conducting observations in a nonthreatening manner. Observer effects also seem to diminish over time.

It should be realized though that to the extent that observer presence affects the classroom behaviors measured and compared, the validity of those comparisons may be reduced.[1]

Population Comparability

Researchers wishing to compare classroom behaviors between countries need to judge the comparability of the populations in each country from which samples might be drawn, and measurements made. Thus, the question of the comparability of samples of behavior needs to be preceeded by the question: what are comparable populations? (Scheuch, 1968, p. 196).

Judgment of the comparability of populations may involve considering questions such as the following. How similar does grade level need to be in order to help ensure that meaningful comparisons can be made? Can subjects given the same labels in different countries be compared, and what should be done if labels differ? Can behaviors occurring in traditional classroom settings be meaningfully compared with behaviors in open classroom situations? Are first year college classes in Nepal, which represent the eleventh year of schooling, meaningfully comparable with first year college classes in the U.S.A., which represent the thirteenth year of formal schooling, or should they be compared with U.S. eleventh grade high school classes? Can teacher and student behaviors occurring in classes containing only two or three students (as was sometimes the case in Nepal) be meaningfully compared with behaviors occurring in classes containing 25 or 30 students?

Although not judged to be a significant problem during the investigation of Nepalese and U.S. classroom behaviors, somewhat serious problems of population comparability might be faced by other researchers wishing to meaningfully compare classroom behaviors cross-nationally--and for this reason, a reminder of this potential problem is given here.

Quality of Sampling Frames

A sampling frame is a list or other device which

serves as the operational definition of the population under study (Warwick & Osherson, 1973, p. 36). The quality of sampling frames will, obviously, affect the representativeness of samples selected. Nonrepresentativeness resulting from the oversampling of some groups and undersampling of others, can result in the introduction of error into inferences later made (Warwick & Osherson, 1973, pp. 37-39).

According to Mitchell (1973), "adequate" sampling frames are often unavailable in many countries. When samples are selected using such frames, sampling biases may result whose exact nature and extent are difficult to determine (Mitchell, 1973, p. 208).

Sampling frame adequacy was found to be somewhat of a problem during the Nepalese survey, especially for one of the grade levels studied. The sampling frame used for that grade level was out-of-date partly due to a national reorganization of schools being carried out which shifted classes at some schools to other nearby schools.

Sample Sizes and Representativeness

Classroom observation recordings are expensive to obtain (Medley & Mitzel, 1963, p. 247)--generally much more expensive in terms of time and money than data gathered using relatively indirect observation techniques such as questionnaires. The expensiveness of direct observation techniques tends to place restrictions upon the sizes of

samples selected, and the representativeness of those samples in terms of national populations. For example, none of the existing data pools descriptive of U.S. classroom behaviors used in the study of Nepalese and U.S. teaching were taken from populations larger than those contained in three cities of one state, and none were based upon the observation of more than about 30 teachers.

Such restrictions place limitations upon the accuracy of inferences which can be drawn about national populations, and limit the accuracy of comparisons which can be made between nations.

Cycles in Classroom Behaviors

Another factor which may affect the representativeness of behaviors observed, and the validity of comparisons later made, concerns cycles of classroom behaviors that may occur over time.

> Interaction patterns cycle through the school year, the semester, the 2- to 4- week unit of study, the day, the daily lesson, and in terms of learning to spell a particular word in a spelling lesson. . . . The relationship between these factors and interaction in the classroom is sufficiently strong to lead to incorrect conclusions if they are ignored. (Flanders, 1970, p. 156)

Thus, care should be taken to ensure that differences in phases of instruction do not bias comparisons made between different groups (Flanders, 1970, p. 156). Unfortunately, little empirical information now exists describing the kinds of cyclical differences which do occur during

instruction.

When the survey of classroom behaviors was conducted in Nepal, about 2/3's of the classes observed were visited from March to June, and about 1/3 were visited from August to October. It was not known then, and still is not known, if and how teaching behaviors vary during the school year in Nepal. If variation does occur to any extent, then the normative results contained in reports of the Nepalese survey (Pfau et al., 1975; Pfau, 1977b) may be less representative than is now assumed.

Encoding Concerns

Introductory Remarks

In addition to the sampling strategies used, the ways in which classroom behaviors are encoded can have an effect upon the validity of comparisons later made. "Encoding", as used here, refers to the conversion of observed behaviors into a form suitable for counting and tabulation (Biddle, 1967, p. 338). Encoding occurs when a behavior is observed and an observer classifies it using a category and records the related code symbol (Flanders, 1970, p. 120).

The validity of comparisons can be reduced if different observers encode behaviors differently. As will be seen, differences in the ways that behaviors are encoded can result from conscious decisions made by researchers,

unconsciously, and as a result of the way that a category system is designed and recordings are made using it.

Intra-Investigation Reliability

A researcher studying the classroom behaviors of different countries may wish to compare existing data gathered by other researchers who used the same category system. Or, he may conduct or coordinate studies aimed at describing classroom behaviors in the countries of interest, and then compare the resultant descriptions. A major consideration of such a researcher should be the "intra-investigation reliability" of the data compared.

Intra-investigation reliability refers to the similarity of usage of the same instrument during different investigations (Rosenshine & Furst, 1973, p. 168). That is, it refers to the question of whether the same observation instrument was used the same way by different investigators, so that data gathered may be validly compared.

The need for similarity of instrument usage if valid comparisons are to be made was summarized by Przeworski and Teune (1970), when discussing direct measurements and comparative research.

> The rules of empirical interpretation of the terms of a language must be applied consistently to all observations. . . . To the extent that the rules are not applied consistently, there is a loss in reliability. To the extent that inconsistent application of the rules tends to be concentrated in one country rather than be distributed across all countries, there is

> a systematic error in observations. Such error must be removed if the findings are to be reliable. (p. 98)

During the study of Nepalese and U.S. classroom behaviors, a number of problems related to intra-investigation reliability were discovered. On the one hand, it was found that researchers often did not use the same category system in exactly the same way. For example, it was discovered that some researchers differed in their use of the FIAC by (a) recording teacher writing on the blackboard or overhead projector by using category 5 (Campbell, 1968; Hastings, 1970), versus the standard practice of recording all types of silence, including that occurring while the teacher writes on the blackboard, by using category 10, by (b) recording all teacher repetition of student answers by using category 3 (Campbell, 1968; Flanders, 1969), versus the practice suggested by Amidon and Flanders (1971, p. 28) and used in the Nepal study, of recording teacher repetition of correct student answers by using category 2, or by (c) using the FIAC continuously, from the start of a class period to the end (Kim, 1972), or by using a tape recorder to record whole class discussions as well as talk with individual students which occurred during laboratory and "seat work" situations (Snider, 1966), versus the practice (used in the Nepal study) of recording only periods of teacher lecture and discussions which could be heard by the entire class.

In addition, it was found that researchers fre-

quently did not present evidence indicating that their use of a category system was standardized. That is, common practice in most projects reviewed was to conduct reliability checks to determine the extent agreement between project observers using a particular category system such as the FIAC. However, only infrequently were efforts made to determine if observers were using the category system in a standard manner, by seeing if their observation agreed with those made by "expert" users of the instrument. Such reliability checks with expert standard users seem to be important if comparisons are to be made between investigations, for it is possible that a team of project observers may reach high levels of agreement among themselves when using an instrument, but their usage may differ from that intended by the instrument's developer.[2]

Observer Drift in Reliability

During the survey of teaching in Nepal, a phenomenon was observed which this investigator called "observer drift". That is, in many cases, observer usage of an instrument was found to differ from standardized usage over time (Pfau, 1977b, pp. 56-60).

Other researchers also seem to have encountered this phenomenon. For example, Ryans (1960) indicated that observer reliability, as measured by correlations of trainee and senior observer assessments at the end of training, was high. However, during his project, this

level of agreement was usually not maintained over an extended period of observation, and refresher training was often required (Ryans, 1960, p. 93). Flanders also touched upon this problem when he stated that:

> The problem of observer training is twofold: first, converting men into machines, and second, keeping them in that condition while they are observing. (1967, p. 158).

The point is that this phenomenon of drift from standardized usage when it occurs, needs to be controlled if maximally reliable (and valid) measurements and comparisons are to be made. Researchers also need to be aware that conducting only one reliability check during the course of an observational study (as is now frequently done) may not be sufficient to judge the reliability of observations made.

Time Unit Distortion

Different category systems make use of different techniques to standardize the frequency or pacing of recording. Karafin (1973) points out that frequently an arbitrary time unit is used, which involves the recording of at least one category symbol at the end of every number of seconds defined as a single unit (e.g., every 3 seconds with the FIAC, and every 5 seconds with the ACI). Other units of measure often used include those which "utilize either every category change, subject area change, or every speaker change as one complete unit"

(Karafin, 1973, pp. 21-22).

No matter what units are used, however, a certain amount of distortion takes place during the process of encoding and decoding observational records. When time units are used, difficulties are encountered because "classroom events may be operating at another rhythm, and encoding is different when the events 'break' on or between the arbitrary time boundaries" (Biddle, 1967, p. 341). Investigators who use other kinds of units obtain different forms of distortion, because they are forced to treat all units as if they were the same length (Rosenshine & Furst, 1973, p. 136).

The ACI has a 5 second time interval as its basic unit. Classification of events using that interval is guided by a number of factors, including a ground rule which states that when two or more activities occur sequentially within any 5 second interval, observers should record the activity which occupied the major portion of the interval. Thus, if a teacher asked a question for 3 seconds and a student gave a 2 second answer during the 5 second time interval, an observer would record only category 7 (teacher questioning). No record would be made indicating that student speaking had occurred.

As a result, distortions of reality can occur, especially if certain activities measured by the instrument tend to be quite short (such as student answers sometimes are). Additionally, if the tempos of classroom behaviors

measured using the instrument tend to differ from country to country, the danger exists that different distortions might occur in different countries. These differing distortions, in turn, might reduce the validity of comparisons made between the countries.

Given the apparent importance of this problem a somewhat detailed discussion, focusing on the FIAC, is presented to further indicate the kinds of distortions which may occur.

Flanders, when he developed the FIAC, combined a 3 second time interval with category changes, depending upon which came first, as his unit of measure (Karafin, 1973, p. 22). This is shown by the FIAC "ground rule" which states that:

> If more than one category occurs during the three-second interval, then all categories used in that interval are recorded; therefore, record each change in category. If no change occurs within three seconds, repeat that category number. (Amidon & Flanders, 1971, p. 27)

When behaviors are recorded, however, the tempo of recording is maintained at as regular a rate as possible--usually every 3 seconds.

If the rhythm of behaviors occurring in the classroom is such that the behaviors last an average of 3 seconds, or multiples of 3 seconds, no distortion is encountered. However, if the rhythm of behaviors is such that some are less than 3 seconds in length, and others are slightly longer, then a recording made using the above-

mentioned ground rule will accurately represent the sequence of behaviors which occurred, but may distort estimates of the time spent in each category of behavior.

Table 3 gives an idea of distortions which might occur if some classroom behaviors differed in tempo from a 3 second rate. As can be seen, if such is the case, quite a few calculated values will differ from their "actual values".

The extent to which such distortions are reflected in the Nepalese and U.S. data compared is unknown. Table 3 is meant only to illustrate distortions which might have taken place, not distortions which actually took place, for the "actual values" shown are based upon somewhat extreme assumptions about the rapidity of Nepalese teacher praise and student responses (but assumptions in keeping with general findings of the Nepalese survey nonetheless).

Undoubtedly, however, time interval distortions did occur in both countries, in different ways and to different degrees, when observations were encoded and later decoded. Undoubtedly, also, the validity of comparisons made between Nepal and the U.S.A. was affected by those distortions. Whether affects were negligible or significant is unknown at this time, and will remain unknown until further research concerning such distortions is conducted.

Use of Different Encoding Tempos

Let us now focus upon only the tempo of encoding

TABLE 3

TIME INTERVAL DISTORTIONS OCCURRING IF TEACHER PRAISE AND STUDENT TALK--RESPONSE LASTED FOR ONLY 1 SECOND AT A TIME

Variable	Calculated Value[a]	"Actual" Value[b]
FIAC category		
1. Accepts feelings	.0%	.0%
2. Praises or encourages . .	6.3%	2.4%
3. Accepts or uses ideas of students	.0%	.0%
4. Asks questions	13.9%	16.0%
5. Lecturing	49.5%	56.9%
6. Giving directions	2.9%	3.3%
7. Criticizing or justifying authority	1.0%	1.1%
8. Student talk--response . .	13.2%	5.1%
9. Student talk--initiation .	2.3%	2.6%
10. Silence or confusion . . .	11.0%	12.7%
Total talk		
Teacher talk	73.5%	79.7%
Student talk	15.4%	7.6%
S/(S+T) ratio	.174	.088

Note. The calculations shown are based upon recordings made in the fifth grade mathematics classes of one Nepalese school.

(a) Based upon standard FIAC calculating procedures, which give equal weightage to each category number recorded.

(b) These idealized values are based on the assumption that teacher praise and student talk--response occurred for only 1 second at a time, whereas other categories of behavior recorded occurred for exactly 3 seconds at a time, or for exact multiples of 3 seconds.

classroom behaviors, and ignore the type of problem discussed in the last section.

When the FIAC is used,

> observation continues at a rate of 20 to 25 tallies per minute, keeping the tempo as steady as possible. This usually works out to about one tally every 3 seconds. There is nothing magical about a 3-second period. An experienced observer, after considerable practice, tends to classify at this rate with this particular category system. A gifted observer might settle down to a faster rate, after considerable experience. . . . (Flanders, 1970, p. 37)

Flanders, however, points out that the same recording tempo should be used if comparisons are to be made.

> Having a regular tempo is much more important than achieving a particular rate because most conclusions depend on rate consistency, not on speed. For example, a comparison between two categories in one observation or the same category in two different observations is possible only when the tempo of coding is the same for both categories and for both observations, whether or not that tempo is one tally every 2, 2½, 3, or 3½ seconds. (1970, p. 37)

Pfau (1977b, pp. 248-251) revealed more specifically the kinds of distortions which can occur when different recording tempos are used.

Such information implies that the validity of comparisons made between countries may be reduced if observers in those countries encode classroom behaviors at different rates using instruments such as the FIAC.[3]

Encoding Subjectivity

Encoding observations using category systems such as the FIAC and ACI is not a completely objective process. Observers using these instruments constantly make judgments

about the behaviors classified. Most of these judgments are guided by relatively clear rules about what should be done. Some judgments, however, are made more subjectively, and as a result, may affect the comparability of observations recorded in different countries.

During the study of Nepalese and U.S. classroom behaviors, several kinds of judgments associated with use of the FIAC appeared to be more subjective than others. These concerned (a) when to start and stop recording using the FIAC, and (b) distinguishing between the FIAC category 8 (student talk--response) and category 9 (student talk--initiation). Persons interested in knowing more about these judgments are referred to Pfau (1977b, pp. 238-241).

Decoding Observer Records

Equivalence of Behavioral Indices

A purpose of decoding is to recreate in easily understandable symbols those aspects of the original classroom behaviors which were encoded (Flanders, 1970, p. 120). This should be done in such a way that interpretations resulting from the decoding process accurately reflect behaviors which took place. Of concern to any comparative researcher should be the equivalence of variables compared between cultures and nations.

In addition to difficulties associated with the absence, irrelevance, or differing distributions of variables in different cultures (Armer, 1973, p. 70), researchers

seeking to describe and compare the classroom behaviors of different cultures may encounter problems associated with (a) similar but functionally different behaviors, and (b) functionally equivalent but different behaviors.

> Perhaps a variety of instructional procedures exists for attaining the same educational goals. These behaviors may differ in overt properties and thus may be coded differently by an observer, but they may engender the same implicit behaviors in students. . . . For example, teacher probing of student responses might be equivalent to the use of student ideas; homework questions requiring higher level processing might be equivalent to the higher level teacher questions in classroom interactions; requiring students to respond at frequent intervals might be equivalent to class discussion and frequent questions; teacher disapproval which is followed by eliciting student clarification of what the student said might be equivalent to teacher use of student ideas. (Rosenshine & Furst, 1973, pp. 172-173)

The functional equivalence of such behaviors is of concern when inferences are made from measurements of those behaviors. That is, it is of concern when inferences from the same kinds of direct measurements to inferred measurements are not equally valid for all systems being studied (Przeworski & Teune, 1970, p. 104).

For example, using the FIAC, all teacher statements are considered to be either "indirect" or "direct" teacher behaviors.[4] According to Amidon and Flanders (1971, p. 6), indirect teacher behaviors tend to maximize the freedom of students to respond, and direct teacher behaviors tend to minimize the freedom of students to respond. FIAC categories 1, 2, 3, and 4 are considered to represent relatively indirect behaviors, and categories 5, 6, and 7

represent direct behaviors. Measurements made using these categories are used to calculate the values of several ratios (called the I/(I+D) and i/(i+d) ratios) whose magnitudes are used to infer the extent of indirect and direct teacher influence.[5]

Although such inferences may be appropriate to U.S. classroom behaviors, this investigator discovered problems of equivalence when the same procedures were used in Nepal. That is, in Nepal, this investigator felt that "teacher questioning" may not represent an indirect behavior as much as a direct one, and "giving directions" may often be considered more an indirect behavior than a direct one--this interpretation being exactly the opposite of that made in the U.S.A! (see Pfau, 1977b, pp. 252-254 for details). As a result, the I(I+D) and the i(i+d) ratios as usually defined may not be valid indicators of Nepalese teacher directness and indirectness, and thus may not be meaningfully compared between the U.S.A. and Nepal.

It might be noted, however, that even though difficulties were encountered with these two ratios, the measurements of most variables studied in Nepal and in the U.S.A. were judged to be such that meaningful comparisons of them could be made.

Problems of Context and Meaning

Observers using category systems such as the FIAC, record only behaviors described by the category system. By

doing so, they and data interpreters essentially ignore much of the meaning of what is said, and the context in which the behaviors took place. In part, the assumption is made that classroom events can be analyzed in terms of small bits of behavior that have the same value irrespective of context (Walker, 1972a, p. 44).

Ignoring the meaning of what is said and its context may result in interpretations during decoding which are misleading. For example, ignoring contextual factors may force the decoder into the situation of not knowing whether the students were or were not paying attention to the teacher, of not knowing whether student "initiation" was cooperative or whether it represented uncooperative resistance to the teacher or other "off-target remarks" by students, and of not knowing whether silence or confusion had taken place (Walker, 1972b, p. 33; Flanders, 1970, p. 50). One scholar, after considering such difficulties, even went so far as to state that the FIAC "is, by its very nature, inadequate as a tool for the analysis of teaching" (Fortgang, 1972, p. 11).

It seems though that many of these difficulties can be dealt with, by first conducting preliminary studies in the cultures to be compared, and by secondly, modifying existing instruments or developing new instruments so that realities of the classroom situation of concern to the investigator will be faithfully recorded and interpreted. For example, based upon preliminary unstructured observa-

tions, if it were thought desirable, the FIAC could be subscripted[6] to indicate behaviors such as whether student "initiation" was cooperative or not, or what portion of "silence and confusion" consisted of nonproductive confusion and what portion consisted of productive silence. Also, if thought necessary in order to reflect other relevant conditions, other instruments could be developed and used to determine, for example, what percentage of students were paying attention to the teacher. By adopting such an approach, perhaps the more serious problems of context and meaning could be adequately dealt with.

The Sequential Analysis of Behaviors

When category systems such as the FIAC are used, it is possible to study not only the frequency occurrence of different categories of behavior, but also sequential relationships between those categories. That is, the records made by using many category systems can be used to study patterns of classroom behaviors in some detail (Rosenshine & Furst, 1973, p. 167).

Matrices have frequently been used to study patterns or sequences of behaviors recorded using the FIAC (e.g., Payne, 1970; Bosch, 1972). Such matrices are constructed by considering a pair of events as the unit to be tabulated, rather than a single event (Flanders, 1970, p. 75).

However, a fundamental difficulty is encountered

when matrices are used to study patterns of behavior longer than two units in length, as is done when the technique called Teaching Pattern Analysis is applied (Amidon & Amidon, 1967). The difficulty is that three, four and five unit patterns cannot validly be constructed from the two unit data represented in matrices (Prokop, 1969, p. 57; Campbell, 1975, p. 264). That is, when extended patterns are derived from matrices, as was done using Teaching Pattern Analysis during the Nepal/U.S. study, the extent to which those patterns represent reality is unknown, and cannot be estimated. The patterns identified do not necessarily represent actual sequences of behaviors which took place.

In short, the validity of inferences and comparisons made based upon Teaching Pattern Analysis may suffer as a result of inherent limitations of the technique.[7]

Failure to Follow Project Protocol

Another potential problem of concern to comparative researchers is the failure of personnel to follow a project's protocol (i.e., planned methodology). Such a problem may occur either unconsciously (as Barber, 1973, p. 390 indicates) or as a result of dishonesty (as Elder, 1973, p. 136 experienced to a minor degree).

During the survey of Nepalese teaching, this problem was encountered in at least three separate ways. Several times, research assistants did not or could not

follow the project's sampling plan; through a misunderstanding, one team of observers recorded observations for only part of some class periods observed (rather than the entire periods as planned); and one temporarily hired assistant was found to be falsely constructing FIAC matrices.

Such failures to follow project protocol represent yet another factor which can reduce the validity of cross-national comparisons.

Discussion

As one can realize from the previous review, quite a few factors may reduce the validity of cross-national comparisons made using category systems. In terms of the Nepalese and U.S. comparisons made by this investigator, the individual and combined effects of these factors are unknown. However, given the positive indications of validity found during that investigation (Pfau, 1977a; 1977b, pp. 213-218), it appears as if category systems remain worthy of consideration by scholars wishing to make relatively precise and valid cross-national comparisons of classroom behaviors. Additional research is obviously needed, however, to further explore strengths and weaknesses associated with category system usage.

Notes

[1]At times it may be helpful to keep in mind that teachers and students cannot do what they cannot do--even if they behave differently (Kerlinger, 1973, p. 539). That is, persons cannot usually act in ways that they have not learned to act. Also, and somewhat pragmatically, although observer effects may occur to some extent, it may sometimes be better to know how teachers and students behave while they are under observation than to know almost nothing about how they behave in the classroom (Medley & Mitzel, 1963, p. 248).

[2]During the study of teaching in Nepal, conscious efforts were made to ensure that the FIAC and ACI were used in standardized ways. The approach used was to first check this investigator's usage of the FIAC with standard observations contained in several training kits containing tape recordings of U.S. classes, and to check his usage of the ACI by observing several science classes together with the ACI's developer, Dr. Harrie Caldwell. Agreement was found to be high. Then, the investigator trained observers in Nepal until their usage of these instruments when observing Nepalese classes was similar to his own usage. It was assumed that Nepalese observers whose use of the instruments was similar to this investigator's usage (in Nepali), were making observations in a manner consistent with standardized usage in the U.S.A. (in

English).

[3]It should be noted however, that users of instruments such as the ACI are not faced with the problem of distortions resulting from the use of different recording tempos, since all ACI observers use the same tempo, a 5 second one, which is standardized by referring to stopwatches.

[4]The FIAC is derived from a research tradition, which includes H.H. Anderson's studies of the integrative and dominative behavior of teachers (1939, 1945, 1946), Lippitt and White's (1943) studies of authoritarian, democratic, and laissez-faire leadership patterns, and Withall's (1949) studies of teacher-centered and learner-centered teaching.

[5]The I(I+D) ratio is equal to the sum of measurements made using categories 1, 2, 3, and 4 divided by the sum of measurements made using categories 5, 6, and 7. The i/(i+d) ratio is equal to the sum of measurements made using categories 1, 2, and 3 divided by the sum of measurements made using categories 6 and 7.

[6]Subscripting means dividing a single category into additional subcategories (Flanders, 1970, p. 126).

[7]Fortunately, if a computer is available, a newer technique called MACROanalysis can be used to study sequences of behaviors. This technique indicates dominant patterns of three, four, or five unit lengths, ranks these

patterns, indicates the percentage occurrence of each pattern, and also indicates the frequency occurrence of each category recorded (Shymansky, 1974; Campbell, 1975; Shymansky et al., 1976). This technique seems to overcome a number of difficulties associated with Teaching Pattern Analysis.

References

Adams, R. S., & Biddle, B. J. Realities of teaching: Explorations with video tape. New York: Holt, Rinehart, & Winston, 1970.

Anderson, H. H. The measurement of domination and of socially integrative behavior in teachers' contacts with children. Child Development, 1939, 10, 73-89.

Anderson, H. H., & Brewer, H. M. Studies of teachers' classroom personalities: Dominative and socially integrative behavior of kindergarten teachers. Stanford, Cal.: Stanford University Press, 1945.

Anderson, H. H., & Brewer, J. E. Studies of teachers' classroom personalities: Effects of dominative and integrative contacts on children's classroom behavior. Stanford, Cal.: Stanford University Press, 1946.

Amidon, E., & Amidon, P. Teaching pattern analysis manual. Minneapolis, Minn.: Association for Productive Teaching, 1967.

Amidon, E. J., & Flanders, N. A. The role of the teacher in the classroom: A manual for understanding and improving teacher classroom behavior (Rev. ed.). Minneapolis, Minn.: Association for Productive Teaching, 1971.

Armer, M. Methodological problems and possibilities in comparative research. In M. Armer & A. D. Grimshaw (Eds.), Comparative social research: Methodological problems and strategies. New York: John Wiley, 1973.

Barber, T. X. Pitfalls in research: Nine investigator and experimenter effects. In R. M. W. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally, 1973.

Biddle, B. J. Methods and concepts in classroom research. Review of Educational Research, 1967, 37(3) 337-357.

Bosch, A. C. Relationships of teaching patterns to indices of classroom verbal interaction behavior: A further analysis and synthesis of classroom verbal interaction data using descriptive indices of behavior and Teaching Pattern Analysis (Doctoral dissertation, New York University, 1972). Dissertation Abstracts International, 1973, 33, 7077A. (University Microfilms No. 72-20, 620)

Caldwell, H. E. Evaluation of an in-service methods course by systematic observation of classroom activities. Project No. 6-8760, U.S. Department of Health, Education, and Welfare, Office of Education, 1967. (ERIC Document Reproduction Service No. ED 024 615)

Caldwell, H. E. Evaluation of an in-service science methods course by systematic observation of classroom activities (Doctoral dissertation, Syracuse University, 1968). Dissertation Abstracts, 1969, 29, 4346A-4347A. (University Microfilms No. 69-7725)

Campbell, J. R. Cognitive and affective process development and its relation to a teacher's interaction ratio: An investigation to determine the relationship between the affective and cognitive development of junior high low achievers and the interaction ratio employed by their instructors (Doctoral dissertation, New York University, 1968). Dissertation Abstracts, 1969, 30, 1069A-1070A. (University Microfilms No. 69-11, 740)

Campbell, J. R. Macroanalysis: A new development for interaction analysis. Journal of Educational Research, 1975, 68(7), 261-269.

Dunkin, M. J., & Biddle, B. J. The study of teaching. New York: Holt, Rinehart, & Winston, 1974.

Elder, J. W. Problems of cross-cultural methodology: Instrumentation and interviewing in India. In M. Armer & A. D. Grimshaw (Eds.), Comparative social research: Methodological problems and strategies. New York: John Wiley, 1973.

Flanders, N. A. The problems of observer training and reliability. In E. J. Amidon & J. B. Hough (Eds.), Interaction analysis: Theory, research and application. Reading, Mass.: Addison-Wesley, 1967.

Flanders, N. A. et al. Teacher influence patterns and pupil achievement in the second, fourth, and sixth grade levels (2 vols.). Ann Arbor: University of Michigan, 1969.

Flanders, N. A. Analyzing teaching behavior. Reading, Mass.: Addison-Wesley, 1970.

Fortgang, S. J. An examination of the assumptions of the Flanders system of interaction analysis from the perspective of symbolic interaction theory (Doctoral dissertation, Syracuse University, 1972). Dissertation Abstracts International, 1973, 34, 1175A-1176A. (University Microfilms No. 73-19, 807)

Hastings, H. I. A study of the relationship between teacher-pupil verbal interaction and pupil achievement in elementary school science (Doctoral dissertation, University of Oregon, 1970). Dissertation Abstracts International, 1971, 31, 5033A-5034A. (University Microfilms No. 71-10, 730)

Karafin, G. R. Discussion of considerations for selecting or developing an observation system. Classroom Interaction Newsletter, 1973, 8(2), 15-32.

Kerlinger, F. N. Foundations of behavioral research (2nd ed.). New York: Holt, Rinehart & Winston, 1973.

Kim, C. S. A preliminary normative study on teaching behavior in elementary school. Seoul: College of Education, Seoul National University, 1972.

Lippitt, R., & White, R. K. The "social climate" of children's groups. In R. G. Barker, J. S. Kounin, & H. F. Wright (Eds.), Child behavior and development. New York: McGraw-Hill, 1943.

Medley, D. M., & Mitzel, H. E. Measuring classroom behaviors by systematic observation. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Mitchell, R. E. Survey materials collected in the developing countries: Sampling, measurement, and interviewing obstacles to intra- and inter-national comparisons. In D. P. Warwick & S. Osherson (Eds.), Comparative research methods. Englewood Cliffs: Prentice-Hall, 1973.

Nuthall, G., & Church, J. Observation systems used with recording media. International Review of Education, 1972, 18(4), 491-507.

Payne, J. I. Analysis of teacher-student classroom interaction in Amish and non-Amish schools using Flanders Interaction Analysis technique (Doctoral dissertation, Pennsylvania State University, 1970). Dissertation Abstracts International, 1971, 32, 251A. (University Microfilms No. 71-16, 650)

Pfau, R. H., Mathema, S. S. B., Pradhan, K. M., Rayamajhi, L. B., & Shrestha, S. K. A national study of classroom behaviors in Nepalese schools: Introduction, methodology, summary of findings, and general recommendations (Vol. 1). Final Report. Kirtipur, Nepal: Research Division, Institute of Education, 1975.

Pfau, R. H. The cross-national comparison of classroom behaviors. Paper presented at the annual meeting of the Comparative and International Education Society, February 1976. (ERIC Document Reproduction Service No. ED 120 260)

Pfau, R. H. A comparison of Nepalese and U.S. classroom behaviors. Paper presented at the annual meeting of the Comparative and International Education Society, February 1977. (ERIC Document Reproduction Service No. ED 137 265) (a)

Pfau, R. H. A cross-national comparison of classroom behaviors: Based upon a survey conducted within Nepal using the Flanders Interaction Analysis Categories and Caldwell's Activity Categories Instrument. Unpublished doctoral dissertation, University of Pittsburgh, 1977. (b)

Prokop, M. Sequential analysis of verbal behavior patterns in a Flanders Interaction Analysis matrix. Classroom Interaction Newsletter. 1969, 4(2), 53-58.

Przeworski, A., & Teune, H. The logic of comparative social inquiry. New York: John Wiley, 1970.

Rosenshine, B., & Furst, N. The use of direct observation to study teaching. In M. W. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally, 1973.

Ryans, D. G. Characteristics of teachers: Their description, comparison, and appraisal. Delhi: Sterling Publishers, 1969. (Originally published by the American Council on Education, 1960.)

Samph, T. Observer effects upon teacher behavior (Doctoral dissertation, University of Michigan, 1968). Dissertation Abstracts, 1969, 29, 2573A-2574A. (University Microfilms No. 69-2382)

Samph, T. Observer effects on teacher verbal classroom behavior. Journal of Educational Psychology, 1976, 68(6), 736-741.

Scheuch, E. K. The cross-cultural use of sample surveys: Problems of comparability. In S. Rokkan (Ed.), Comparative research across cultures and nations. The Hague: Mouton & Ecole Pratiques de Hautes Etudes, 1968.

Shymansky, J. A. A comparison of macroanalytic techniques in the study of teacher behavior patterns in the classroom. Classroom Interaction Newsletter, 1974, 10(1), 51-56.

Shymansky, J. A., Penick, J., Good, R. G., & Matthews, C. C. Using macroanalytic techniques to study teacher behavior patterns. Journal of Research in Science Teaching, 1975, 12, 221-227.

Snider, R. M. A project to study the nature of physics teaching using the Flanders method of interaction analysis (Doctoral dissertation, Cornell University, 1966). Dissertation Abstracts, 1966, 26, 7183A-7184A. (University Microfilms No. 66-5078)

Standards for educational & psychological tests. Washington, D. C.: American Psychological Association, 1974.

Walker, R. Some general problems that arise when interaction analysis is used to assess the impact of educational innovation. Classroom Interaction Newsletter, 1972, 7(2), 38-48. (a)

Walker, R. The sociology of education and life in school classrooms. International Review of Education, 1972, 18(1), 32-41. (b)

Warwick, D. P., & Osherson, S. (Eds.), Comparative research methods. Englewood Cliffs: Prentice-Hall, 1973.

Withall, J. G. The development of a technique for the measurement of social-emotional climate in classrooms. Journal of Experimental Education, 1949, 17, 347-361.